DOCUMENT RESUME

ED 081 844 TM 003 205

AUTHOR Pyrczak, Fred
TITLE Use of Similarities between Stems and Keyed Choices in Multiple-Choice Items.
PUB DATE 73
NOTE 10p.; Paper presented at Annual Meeting of National Council for Measurement in Education (New Orleans, Louisiana, February 26-28, 1973)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Item Analysis; *Multiple Choice Tests; Technical Reports; *Test Construction; Testing; Test Reliability; Test Results; Test Validity; *Test Wiseness

ABSTRACT

The general purpose of this study was to determine the effects of similarities between stems and keyed choices on test difficulty. Unlike previous investigations of this undesirable characteristic of some multiple-choice items, the present study employed items that were unintentionally faulty and samples of examinees who were highly experienced test-takers. The first phase of the study indicated that the presence or absence of similarities did not significantly affect examinees' scores. The second phase, however, indicated that examinees easily may be trained to recognize and use such similarities to advantage. (Author)

FILMED FROM BEST AVAILABLE COPY

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRO
DUCED EXACTLY AS RECEIVED FROM
THE PERSON OR ORGANIZATION ORIGIN
ATING IT. POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRE
SENT OFFICIAL NATIONAL INSTITUTE OF
EDUCATION POSITION OR POLICY

ED 081844

TM 003 205

Use of Similarities between Stems and Keyed Choices in Multiple-Choice Items

Fred Pyrczak
California State University, Los Angeles

ABSTRACT

The general purpose of this study was to determine the effects of similarities between stems and keyed choices on test difficulty. Unlike previous investigations of this undesirable characteristic of some multiple-choice items, the present study employed items that were unintentionally faulty and samples of examinees who were highly experienced test-takers. The first phase of the study indicated that the presence or absence of similarities did not significantly affect examinees' scores. The second phase, however, indicated that examinees easily may be trained to recognize and use such similarities to advantage.

Use of Similarities between Stems and Keyed Choices in Multiple-Choice Items*

Fred Pyrczak
California State University, Los Angeles

A recent survey of the literature indicated that nine experts suggest that item-writers should eliminate superficial similarities between the stems and keyed choices in multiple-choice items (Masonis, 1970). The most common type of similarity is a result of repeating some portion of the stem in the keyed choice but not in the distracters. Such similarities, it has been reasoned, may help uninformed examinees select keyed choices.

The effects of violating this item-writing principle and thus creating "faulty" test items have been investigated by means of two approaches. First, in studies of "test-wiseness," items have been written about fictitious material so that there was no correct response from a scholarly point of view. In each of the items, one choice that was similar to its respective stem was included. Examinees in these studies (e.g., Diamond and Evans, 1972; Slakter, 1970a,b) were awarded a point for "test-wiseness" each time they selected a choice that was similar to the stem. In general, it has been found that at least some examinees tend to mark a greater number of such choices than would be expected to mark them on the basis of random guessing alone. Specifically,

* Presented at the 1973 Annual Meeting of the National Council on Measurement in Education.

subjects with higher intelligence-test scores (Diamond and Evans, 1972) and at higher grade levels (Slakter, 1970a) tended to select a larger number of such choices. Furthermore, Slakter (1970b) found that examinees can be trained successfully to select these choices in response to the special types of items used in these studies. The artificial nature of the special items, however, makes it difficult to generalize to examinee behavior in real test situations. Specifically, the items used in the studies cited above differed from real test items in two potentially important respects. First, similarities were intentionally incorporated in the special items used in these studies and thus may be different in nature and degree from those that unintentionally occur in test items. Secondly, the items used in these studies had no "correct" response since the items dealt with fictitious material. Thus, partial knowledge could not be used by examinees when responding to the items. In real test situations, examinees often are able to use partial information (and, of course, adequate information) to determine their responses. Thus, the studies cited above do not indicate whether examinees rely upon superficial clues when they have some knowledge of the points in question.

Using a different approach, McMorris et al. (1972), as part of a larger study, incorporated similarities in seven items of reasonable difficulty. When the original items and the modified items were administered to different groups of examinees, a small but statistically

significant difference was found, with the faulty items being easier. The presence or absence of similarities did not significantly affect either reliability or validity.

This study was conducted in two phases. The methodology of both phases was similar to that used by McMorris et al. It was different from all the previously mentioned investigations, however, in two important respects. First, the faulty test items employed in this study were obtained from a published test. As noted above, in previous studies, similarities were knowingly incorporated in items by the investigators. Thus, a higher degree of generalization can be obtained from this study regarding the effects of faults that unintentionally occur in test items. Secondly, the examinees in this study apparently were much more experienced as test-takers, and thus should exhibit a higher degree of test-wiseness than the examinees in previous studies.

PHASE I

Method

Subjects

The sample consisted of 118 undergraduate and graduate students enrolled in courses in educational research and measurement. Many of them had had experience writing tests as well as extensive experience taking them. They were tested, however, prior to specific discussions in their research and measurement courses of topics relevant to the current investigation.

Instrumentation

Ten vocabulary items that appeared to be faulty with respect to the item-writing principle under consideration were identified in a published reading test appropriate for use with adults. Specifically, in four of the items the first syllables of the stimulus words were identical with the first syllables of the keyed choices (e.g., "pre-"); in four of the items the last syllables were identical (e.g., "-some"); and in two of the items the last three letters were identical (i.e., "-ous"). Only one of the 40 distracters in these items, furthermore, contained such a similarity with its respective stimulus word. This was true despite the fact that it has been suggested that distracters in vocabulary items that have superficial similarities with their respective stimulus words may be effective in attracting examinees who do not know the meanings of the words in question (e.g., Kelly, 1937; Pyrczak, 1971).

A second ten-item form of the test was constructed by changing the keyed choices in the ten original items so that the superficial similarities between the stems and the keyed choices were eliminated. In this process, care was taken to obtain keyed choices that were thoroughly adequate and that were at about the same level of readability as the original keyed choices. These considerations precluded the use of one additional faulty item that had been identified in the published test since an adequate substitute for the keyed choice could not be written.

was designed to determine whether similar examinees can be trained to take advantage of the fault under investigation.

Method

Subjects

The sample consisted of 76 undergraduate and graduate students enrolled in courses in educational research and measurement.

Instrumentation

Two forms of a vocabulary test were constructed. Both contained the faulty vocabulary items identified in the published reading test. The only difference between the two forms was in the directions. For one form, the directions merely indicated that the examinees should "...circle the letter of the choice that most nearly means the same thing as the underlined word." For the other form the directions also indicated that when in doubt about the meaning of a given word, it would pay to select a choice with a similar beginning or ending as the underlined word.

Procedures

A random half of the subjects was administered the form of the test with the standard directions. The other half was administered the form with special "test-wiseness" directions.

While it would have been desirable to have used a larger number of items, the advantage of using real items in which faults were unintentionally present seemed to outweigh the disadvantage of using a small number of items. It is interesting to note, furthermore, that the number of items with this particular fault was greater in the present study than in the studies cited previously.

Procedures

A random half of the subjects was administered the form consisting of the ten faulty items as they originally appeared in the published test. The other half was administered the form of the test consisting of the same ten items in their fault-free form.

Results

The mean and standard deviation for the form of the test consisting of the original items were 7.66 and 1.50, respectively. On the fault-free form they were 7.36 and 1.91, respectively. The difference between the means was not statistically significant at the .05 level.

PHASE II

The negative results of the first phase of this study suggest that the presence or absence of similarities between the stems and keyed choices in vocabulary items does not affect test difficulty even for highly experienced examinees. The second phase of the investigation

Results

The mean and standard deviation for the form with standard directions were 7.63 and 1.44, respectively. For the form with the "test-wiseness" directions, they were 8.53 and 2.02, respectively. The difference between the means was statistically significant at the .01 level.

Discussion

Phase I of this investigation indicates that the presence or absence of superficial similarities between keyed choices and their respective stems does not affect the difficulty of vocabulary tests for examinees who have not had special training. This finding is especially interesting in light of the fact that highly experienced examinees were used in this investigation. Those attempting the faulty items, furthermore, marked correct choices to enough items that it theoretically was possible for them to develop a cue-using strategy of marking choices with similarities when they did not know the meaning of a given stimulus word. The second phase of this study, however, indicates that examinees easily can be trained to take advantage of the fault under investigation.

The implication of the first phase of this study for item writers is that the item-writing principle under consideration may be of little practical importance when writing vocabulary items for examinees who

have not had special training. Since the second phase of this study indicates that examinees easily may be trained to take advantage of violations of the "similarity principle," however, it seems desirable for item-writers to avoid writing keyed choices that are physically similar to their respective stems. As noted above, however, it is not always possible to provide a thoroughly adequate keyed choice without violating this principle. When this is the case with respect to a given item, it seems desirable for item-writers to consider the possibility of providing plausible distracters that also are similar to the stimulus word.

REFERENCES

Diamond, J. J. and [illegible], W. J. An investigation of the cognitive correlates of test-wiseness. Journal of Educational Measurement, 1972, 9, 145-150.

Kelly, V. H. An experiment with multiple-choice vocabulary tests constructed by two different procedures. Journal of Experimental Education, 1937, 5, 249-250.

McMorris, R. F. et al. Effects of violating item construction principles. Unpublished manuscript, Syracuse University, 1972.

Masonis, E. J. Comparing two patterns of instruction for teaching item writing theory and skills. Unpublished doctoral dissertation, University of Pennsylvania, 1970.

Pyrczak, F. Relative attractiveness of selected types of distracters for multiple-choice vocabulary items. Paper presented at the meeting of the National Council on Measurement in Education, New York, February 1971.

Slakter, M. J. et al. Grade level, sex and selected aspects of test-wiseness. Journal of Educational Measurement, 1970a, 7, 119-122.

Slakter, M. J. et al. Learning test-wiseness by programmed texts. Journal of Educational Measurement, 1970b, 7, 247-254.